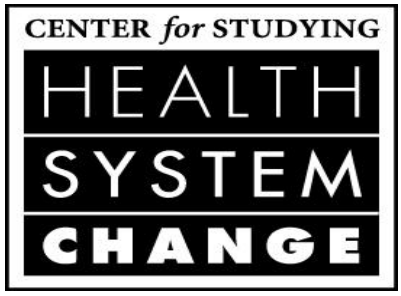
HSC 2008 Health Tracking Physician Survey Methodology Report



Richard Strouse
Frank Potter
Terisa Davis
John Hall
Stephen Williams
Ellen Herbold
Judy Walsh
Ellyn Boukus
Jim Reschovsky

600 Maryland Avenue, SW
Suite 550
Washington, DC 20024
www.hschange.org

Technical Publication No.

77

September 2009

AUTHORS

Richard Strouse, Independent Consultant

Frank Potter, Mathematica Policy Research, Inc.

Terisa Davis, Westat

John Hall, Mathematica Policy Research, Inc.

Stephen Williams, Mathematica Policy Research, Inc.

Ellen Herbold, Westat

Judy Walsh, Westat

Ellyn Boukus, Center for Studying Health System Change

Jim Reschovsky, Center for Studying Health System Change

This is one in a series of technical documents that have been done as part of the Community Tracking Study being conducted by the Center for Studying Health System Change (HSC), which is funded primarily by The Robert Wood Johnson Foundation and is affiliated with Mathematica Policy Research, Inc.

CONTENTS

CONTENTS *(continued)*

Chapter **Page**

CONTENTS *(continued)*

Chapter **Page**

TABLES

EXHIBIT

Figure **Page**

I. OVERVIEW

A. OBJECTIVES OF THE COMMUNITY TRACKING STUDY AND THE HSC HEALTH TRACKING SURVEYS

The Community Tracking Study (CTS) has been the core research effort of the Center for Studying Health System Change (HSC), a nonpartisan policy research organization in Washington, DC, that is funded in part by the Robert Wood Johnson Foundation (RWJF) and is affiliated with Mathematica Policy Research, Inc. HSC's mission is to inform health care decision makers about changes in the health care system at the local and national levels, as well as how such changes will affect people. Since 1995, HSC has conducted five rounds of household and physician surveys; an employer survey was conducted for the first round but discontinued for subsequent rounds.

The first four rounds of CTS surveys were focused on 60 nationally representative communities stratified by region, community size, and whether metropolitan or nonmetropolitan. In addition, the CTS examined 12 of the 60 communities in depth by conducting site visits and using survey samples large enough to draw conclusions about health system change in each community. The 12 communities make up a randomly selected subset of sites that are metropolitan areas with more than 200,000 people (as of July 1992). For budgetary reasons, the community-based design was replaced by a national sample design for the 2008 Health Tracking Household and Physician Surveys, although site visits continue to be focused on the 12 communities (6 rounds of site visits have been completed, with the latest occurring in 2007). Because the 2008 samples are no longer clustered in communities, the surveys have been renamed the HSC Health Tracking Household and Physician surveys. The name change

occurred after the field period commenced, so the survey documents contained in the Appendices still refer to the 2008 physician survey as the CTS physician survey.

The original plan for the 2008 survey was a dual-mode survey, using computer-assisted telephone interviewing (CATI) and self-administered mail questionnaire. This was motivated by the desire to transition to a self-administered mail survey and to maintain the ability to track changes from previous rounds of the CTS surveys. A 2006 pilot study conducted by HSC found that there were substantial mode effects between CATI and mail questionnaire responses. A dual-sample, dual-mode survey would allow tracking from previous rounds by using statistical adjustments. Data collection for the CATI portion of survey commenced prior to the mail portion. From the start, it became apparent that CATI response rates would be low and difficulties contacting physicians would unacceptably increase costs for CATI data collection, reducing funds available for the mail survey. Consequently, the CATI portion of the survey was abandoned and the survey became a single-sample, single-mode, self-administered mail survey. Discussion of the CATI portion of the survey is not included in this report.

B. THE 2008 HEALTH TRACKING PHYSICIAN SURVEY

For each round of the CTS physician surveys and the 2008 Health Tracking Physician survey, a sample of practicing physicians across the country offers perspective on how health care delivery is changing. For each of the first three rounds, more than 12,000 physicians were interviewed by telephone. The number of telephone interviews was reduced to approximately 6,600 physicians for Round Four. In 2008, the community-based sample design was replaced by a nationally-representative sample and the method of data collection was changed from telephone to mail. These changes reflected increased difficulty over time in persuading

physicians to participate in telephone-administered surveys, which raised the cost of conducting surveys and reduced survey response rates. Moreover, limited funding called for a more efficient national sample rather than one clustered in communities. The use of a national sample no longer allows for estimates at the individual community level, but national estimates can be made using smaller samples while maintaining precision.

A study conducted prior to the 2008 survey indicated that many of the variables tracked from earlier rounds were likely to be affected by the change in survey mode and that a shift from telephone to mail data collection with a pre-paid incentive was likely to increase the response rate.[1]

In the 2008 survey, a total of 4,720 physicians replied to the mail survey and the weighted response rate was 61.9 percent. Physicians responded to questions on whether they can provide needed services for patients, how they are compensated, the impact of care management strategies on their practices, and their practice arrangements. For the first four CTS rounds, the Gallup Organization conducted the telephone interviewing for the physician survey while MPR was responsible for the sample design, sample weights, variance estimation, and, for rounds two through four, tracing of physicians who could not be located. For the 2008 survey, Westat conducted the mail survey and tracing activities and MPR was responsible for sample design and sampling weights. MPR and Social and Scientific Systems (SSS) collaborated with Westat and HSC to prepare the documentation for the public and restricted use files. Additional background on CTS is available at HSC's website (http://www.hschange.org/).

[1] See HSC Technical Publication No. 71- http://www.hschange.com/CONTENT/889/.

This report describes the survey design and data collection procedures used for the 2008 Health Tracking Physician Survey. We discuss the sample design in Chapter II, instrument design, cognitive interviewing, and survey preparation in Chapter III, data collection procedures in Chapter IV, and sample weighting in Chapter V. Cognitive interviewing protocols are included in Appendix A and the survey instrument and advance materials mailed to physicians in Appendix B. Reports describing the first four rounds of the CTS physician survey are included in Technical Publications #9, #32, #38, and #70.

II. INSTRUMENT DESIGN AND COGNITIVE INTERVIEWING

The survey instrument was designed by a team of HSC staff, in consultation with Robert Wood Johnson Foundation staff and other experts. In light of the original plan for a two-sample, dual-mode self-administered mail and CATI survey, a number of questions in the 2008 survey instrument were either copied or adapted from questions from previous rounds to allow for tracking. However, a substantial portion of the instrument consisted of new questions, both original and modified versions from other surveys. New topic areas included time spent communicating with patients via email or telephone, use of interpreter services, expanded questions regarding health information technologies, receipt of quality and other reports, care management, coordination of care, medical equipment and hospital ownership, malpractice, and receipt of honoraria and gifts from medically related companies.

New questions developed by HSC underwent cognitive testing, conducted by HSC consultant Carolyn Miller. The sample for the cognitive interviews was drawn from the Round Four CTS Physician Survey respondents, stratified by physician's practice type and specialty designation (PCP or specialist), as appearing in the Round Four data. Upon agreeing to participate in the interview, each respondent was sent a mail questionnaire with the new questions and room for comments on content, format, and layout. Questions covered practice organization and ownership, time allocation and reimbursement for communicating with patients, information technology, quality and coordination of patient care, sources of practice revenue and financial interest in medical equipment or hospitals, compensation method, and practice or hospital location information. After completing the survey, the respondent was

contracted by phone for a scheduled, follow-up telephone cognitive interview performed by the consultant.

Twenty-four interviews were completed, representing six solo- or two-physician practitioners, seven working in group practice, five in hospitals, three in HMOs, and one apiece in a community health center, medical school, and "other" setting. The interviewees consisted of 11 specialists, 10 PCPs, and three who identified themselves as both a PCP and a specialist. On average, interviews lasted for 36 minutes and respondents were offered $100 honoraria for completing the cognitive interviews. The final reports on cognitive interviewing are included in Appendix A.

III. SAMPLE DESIGN

For the first three rounds of the CTS Physician Survey, interviews were administered by CATI to a stratified random sample of physicians in the 60 CTS sites, and to an independent, national sample of physicians, referred to as the "national supplement." In Round Four (2004-2005), the national supplement was eliminated and the sample was re-allocated among the 60 sites to obtain a more efficient, proportional and national sample of physicians.[2]

For the 2008 Health Tracking Physician Survey, a stratified random sampling design (similar to the earlier national supplement sample) was used and the site-base sample was dropped. The survey was administered using a self-administered mail questionnaire instead of CATI.

In the following sections, we describe:

- The target population
- Design issues
- Sample size and precision
- Implementation of the sample design (including sample allocation, selection procedures and sample release procedures)

[2] In the first three rounds, target sample sizes were assigned to each CTS site to support site-level estimates (approximately 400 physicians in each of the twelve high-intensity sites and approximately 100 physicians in each of the other 48 sites). In round four, the target sample sizes for each site were assigned in approximate proportion to the weighted number of physicians in the site. The allocation of the target sample size is statistically more efficient (smaller sample size can obtain comparable standard errors for estimates by reducing the variation in the sampling weights) than the allocation for the prior rounds. The allocation in the 2008 survey was independent of the sites and was based on a proportional stratified sample to the 50 states.

A. TARGET POPULATION

The target population was based on information provided on the AMA Masterfile (which includes both AMA members and nonmembers). The AMA Masterfile includes licensed allopathic physicians and osteopathic physicians who obtained graduate training in allopathic medical schools or were identified on state licensing boards. The AMA Masterfile contains the majority of osteopathic physicians listed in the American Osteopathic Association (AOA) listing of osteopathic physicians. In the four prior rounds of the CTS surveys, the frame included physicians from both the AMA and AOA Masterfiles, ensuring coverage of all osteopathic physicians. However, only 0.5% of sampled physicians were listed in the AOA Masterfile while omitted from the AMA Masterfile. To reduce costs associated with acquiring and processing the AOA Masterfile in the 2008 survey, we sampled only from the AMA Masterfile; thus the survey coverage includes only osteopathic physicians who were in the AMA Masterfile.[3]

To meet the initial eligibility criteria for sampling, physicians in the frame must have 1) completed their medical training, 2) practiced within the 50 states and the District of Columbia, and 3) provided direct patient care for at least 20 hours per week. Residents, interns, and fellows were considered to be still in training and were excluded from the sample. The direct patient care criterion resulted in the exclusion of inactive or retired physicians and physicians who were not based in offices or hospitals (e.g. teachers, administrators, and researchers).

The following types of physicians were designated as ineligible for this survey and were removed from the frame:

[3] Based on a comparison of the 2003 AOA Masterfile and the 2003 AMA Masterfile, approximately 85 percent of the osteopathic physicians in the AOA Masterfile were in the AMA Masterfile.

- Specialists in fields that do not focus primarily on direct patient care (see Table III.1);
- Federal employees;
- Graduates of foreign medical schools who are licensed to practice in the United States only temporarily.

Eligible physicians were then classified as either primary care physicians (PCPs) or specialists. PCPs were defined as physicians with a primary specialty of family practice, general practice, general internal medicine, internal medicine/pediatrics, or general pediatrics. All others with survey-eligible specialties were classified as specialists.

TABLE III.1

SPECIALTIES EXCLUDED FROM THE SAMPLING FRAME,
BASED ON AMA MASTERFILE

Allergy and Immunology/ Clinical Laboratory	Epidemiology	Pain Management
Aerospace	Forensic Pathology	Pathology
Anatomic/Clinical Pathology	Forensic Psychiatry	Pediatric Anesthesiology
Anesthesiology	Hematology/Pathology	Pediatric Radiology
Bloodbanking/Transfusion Medicine	Musculoskeletal Radiology	Public Health and General Preventive Medicine
Chemical Pathology	Medical Management	Radiology
Clinical Biochemical Gene	Medical Microbiology	Underseas Medicine
Clinical Pharmacology	Medical Toxicology	Vascular and Interventional Radiology
Cytopathology	Neuropathology	
	Neuroradiology	
	Nuclear Medicine	

B. DESIGN ISSUES

The survey is based on a classical stratified design with proportional allocation. The 2008 survey design is simpler than those from prior rounds, and the key issue was to meet a cost constraint by reducing sample size while achieving the best possible precision for national estimates.

C. IMPLEMENTATION

1. Sampling Frame

The sampling frame was derived from physician records maintained by the AMA. This file contained the most current information available as of July 2007, just prior to the date of the

2008 Health Tracking Physician Survey sample selection. HSC requested the AMA Masterfile vendor to exclude physicians who resided outside of the 50 states and the District of Columbia, who were employed at a federal hospital (including military, US Public Service and Veterans Administration hospitals) or who were retired, inactive or deceased. The AMA Masterfile vendor was directed to include all physicians whose data record indicated "undeliverable" or "do not contact." The AMA statistical Masterfile list provided to HSC contained information on 735,378 physicians. Data fields on the records in the statistical Masterfile included date of birth, specialty, and other information useful for sampling and weight computations. The statistical file did not contain the physician's name, address or telephone number. After the sample was selected, contact information was obtained from the AMA Masterfile vendor only for physicians included in the sample.

The four steps used to construct the frame were:

1. Specify file content and format for ordering the files
2. Verify file content after receiving the AMA files
3. Exclude ineligible physicians
4. Classify records by the sampling stratum (physician classification and region).

After reviewing frequency counts for key items to ensure file accuracy and completeness, physicians who had an ineligible specialty and physicians for whom no information was available for the state of residence (either for office or preferred mailing address) were excluded from the sampling frame. The final sampling frame included 550,260 physicians. Each eligible physician was linked to the appropriate geographic stratum, based on the physician's preferred mailing address from the AMA files. Finally, each physician was classified as either PCP or specialist according to specialty codes from the AMA data files.

2. Sampling Units and Stratification

Stratification, a feature of most large-scale surveys, performs several important functions. Using strata to define populations that are expected to have similar responses can increase survey precision. Another key function of stratification is to ensure an adequate sample size for important study populations. Stratification also helps to achieve optimum allocation for surveys in which some groups exhibit more variability in responses or are more costly to survey than others. The design for the 2008 physician survey used stratification to improve precision, to ensure adequate representation by region, and to control precision for survey estimates of PCPs and specialists.

The population for the sample included physicians in the 50 states and the District of Columbia. The states were divided into 10 geographic strata. The strata were defined to match those used in the four rounds of the CTS physician survey (with the addition of Alaska and Hawaii in one stratum), and were used in prior physician surveys conducted for the AMA. The geographic regions are defined as follows:

1. Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
2. New York
3. Delaware, New Jersey, Pennsylvania, West Virginia
4. District of Columbia, Georgia, Maryland, North Carolina, South Carolina, Virginia
5. Alabama, Florida, Kentucky, Mississippi, Tennessee
6. Arkansas, Louisiana, Missouri, Oklahoma, Texas
7. Indiana, Michigan, Ohio
8. Illinois, Iowa, Minnesota, Wisconsin
9. Arizona, Colorado, Idaho, Kansas, Montana, Nebraska, Nevada, New Mexico, North Dakota, South Dakota, Oregon, Utah, Wyoming, Washington
10. Alaska, California, Hawaii

The 20 sampling strata were formed by crossing the ten regions by whether the physician was classified as a PCP or a specialist. The resulting frame counts are listed in Table III.2.

3. Sample Allocation

For this survey, the goal of the sample allocation was to achieve the highest possible precision for national estimates. The design was based on a proportional allocation of the sample to PCPs and specialists and across regions.

TABLE III.2
FRAME COUNTS FOR THE 2008 HEALTH TRACKING PHYSICIAN SURVEY

AMA Region	Total	PCP	Specialist
Total	550,260	227,921	322,339
1	35,097	13,782	21,315
2	46,029	17,237	28,792
3	50,435	20,201	30,234
4	66,497	26,684	39,813
5	61,667	25,211	36,456
6	60,476	25,113	35,363
7	49,376	21,454	27,922
8	47,689	21,291	26,398
9	66,542	28,988	37,554
10	66,452	27,960	38,492

Source: MPR computation using AMA Masterfile.

4. Sample Selection and Sample Waves

The physician sample was selected in two phases: an initial screening sample and a mail survey questionnaire sample. In addition, a larger augmented sample was selected to provide a reserve for expanding the sample size in any stratum efficiently. The augmented sample included 20,316 physicians across the 20 sampling strata. In the sample selection, we imposed

implicit stratification within the explicit strata using gender, age, practice type (office-base, hospital-based or locum tenens) and zip code of the preferred address.[4] This augmented sample was randomly partitioned into subsamples (called waves) within each stratum. The initial screening sample of 13,551 physicians was chosen by selecting a number of waves. This sample was sent to the AMA Masterfile vendor and the name and address information was returned for 13,135 physicians. We did not receive the name and address information for 426 physicians. The AMA Masterfile vendor does not release name and address information for physicians that specified a "do not contact" status to AMA.

[4] Two forms of stratification were used in the physician survey: explicit and implicit stratification. For explicit stratification, separate subpopulations (strata) were formed and a specific sample size was assigned to each explicit stratum. The sample selection was performed independently in each explicit stratum. Implicit stratification was performed within each explicit stratum by sorting the sampling frame within the stratum by a set of characteristics of the physician (gender, age, practice type and zip code of the preferred address) and a sequential selection procedure was used to select the sample. This process can achieve a sample allocation that is approximately proportional across the variables used in the sorting step.

IV. SAMPLE FILE PREPARATION AND TRACING

A SAS file containing 13,135 records for the mail sample was delivered by MPR to Westat on October 2, 2007. A total of 150 records on this file were for doctors who had already been chosen by HSC for a telephone survey sample which was subsequently dropped from the study (see section I.A). These records were removed, leaving 12,985 records in the mail screening sample. Of these 12,985, 7.6 percent (991 physicians) were ineligible (based on Westat's initial screening contact). A first release of 8,367 physicians was selected from the remaining 11,994 cases and was sent to Westat for the mail questionnaire administration. In this first release, 5,554 records were used to conduct an initial experiment testing the optimal incentive level and follow-up procedure, (see chapter V for further discussion). For the second release, 108 of 150 physicians selected for the CATI survey were added back in, and 690 new eligible cases were added to the mail survey sample, resulting in the final mail sample worked by Westat of 9,165 physicians.

In addition, 306 physicians selected for the sample were designated as "do-not-contact" on the AMA Masterfile and were not included in screening or mail data collection, although they were included in the final sample as non-responses. Seven hundred and seventy nine cases were identified as ineligible based on the initial screening by Westat. The full mail survey sample included 10,250 physicians (9,165 physicians released to Westat, 799 physicians who were initially screened as ineligible by Westat and 306 physicians who were classified as "do not contact" cases by AMA). Ultimately, a total of 3,301 that were selected for the screening sample were never released for the mail survey. The total breakdown of the sample is shown in Table IV.1.

TABLE IV.1.

SAMPLE BREAKDOWN

	Sample size
Release 1 - Experiment	5,554
Release 1 - Remainder	2,813
Release 2	798
Subtotal: cases sent to Westat for mail survey	9,165
AMA No Contact Cases	306
Ineligible based on initial screening contact	779
Total Mail Survey Sample	10,250
Screening sample not released for the full study (includes 110 AMA no contact cases)	3,301
Full Screening Sample	13,551

A. DATA PREPARATION AND TRACING

For some cases, the original file contained both the AMA mailing address and the office address. A new file with a single address was created for these cases: if the office address was different from the mailing address, the office address was used. Address data were reformatted to standardize address field placement and were reviewed by data management staff to check for missing values or formatting errors. Records without phone numbers were then put through a search (using a vendor's services) to obtain phone numbers.

Data were prepared in the format required by the Westat telephone research center (TRC) for conducting the screening calls. Two files were sent to the TRC: one for those with no phone number, which required tracing (1,980 cases), and another for those with a phone number that could be assigned to screening (11,113 cases). Information gathered during tracing included updated addresses and telephone numbers. Using the initial contact information (physician

name, last known office address, year of birth, year of graduation from medical school, and specialty), tracing was initiated on October 18, 2007.

The tracing protocol was based on the results from a previous physician study which indicated the most cost-effective and time-efficient tracing methods. The majority of cases were initially traced using the Internet (Google, Medicare Physician Directory, and Choice Trust), before relying on paid sources such as Directory Assistance and Lexus Nexus. Occasionally, physicians were located through name changes (usually females), medical centers, hospitals, group practices or HMO's.

If the address from the initial contact information matched the one provided by the tracing source, and the latter included a telephone number, the case was updated in the study management system (SMS). If conflicting contact information was identified, the case was screened to determine which information was correct before entering the information into the SMS. If no information was found, the case was coded as non-locatable.

The tracing of the 1,980 cases with initially missing phone numbers was largely successful. As shown in Table IV.2, approximately 68 percent of these cases were successfully traced. The cases that were not successfully traced were assigned a final status of non-locatable.

TABLE IV.2

CASES THAT HAD NO TELEPHONE NUMBER PRIOR TO SCREENING

	Cases with no phone number	Cases for which phone number was obtained	Cases successfully screened or located*
Release 1 - Experiment	1,107	1,047	651
Release 1 - Remainder	172	169	172
Release 2	99	95	65
Not used	602	565	464
Total	1,980	1,876	1,352

* These cases were finalized with a screener code of anything other than "non-locatable."

B. SCREENING

Screening was conducted on all cases to determine whether the selected physician was eligible to participate in the study. Screening began on October 29, 2007, and was complete by February 4, 2008. Although screening calls were attempted for all 13,093 cases, not every case was successfully screened. The screener could not be completed in some cases due to:

- an inability to obtain cooperation from office staff within the maximum call limit of seven (Max Call);
- no answer by a live person at the telephone number, but the physician's name was confirmed by the voicemail message (Name confirmed);
- office staff stating a refusal to participate in the study (Refusal by Office);
- the physician stating a refusal to participate in the study (Refusal by Physician); and

- an inability to find a valid telephone number for the case (Non-locatable).

Study eligibility was determined for cases where screening was successful by speaking directly with the physician or with a knowledgeable person in the physician's office.

Physicians who were deemed ineligible were marked as such in the SMS and did not receive any study mailings. Cases where the physician personally refused were also not included in any additional study activities. Cases with other screener codes progressed to the mailing stage of the study, including those designated as Max Call, Name Confirmed, Refusal by Office, and Non-locatable. Mailings were sent to all of these cases even though they did not have complete screeners.

All cases were given a final screening code prior to the end of tracing and screening procedures. Table IV.3 summarizes the final screener codes.

TABLE IV.3.

SCREENER RESULT CODES

		Screener result codes							Total sample size
		Eligible	Ineligible	Maximum calls	Name confirmed	Refusal by office	Refusal by physician	Non-locatable	
Release 1 - Experiment	Arm 1	1,527	92	38	69	23	22	283	2,054
	Arm 2	642	48	0	0	0	12	0	702
	Arm 3	610	45	0	0	0	15	0	670
	Arm 2a	788	50	23	35	9	12	157	1,074
	Arm 3a	773	50	18	32	13	23	145	1,054
	Total experiment	4,340	285	79	136	45	84	585	5,554
Release 1 - Remainder		1,783	107	85	130	59	36	613	2,813
Release 2		584	35	18	24	10	12	115	798
Remaining sample (not released)		2,202	1,126	49	84	29	30	408	3,928
Total		8,909	1,553	231	374	143	162	1,721	13,093

V. CONDUCT OF THE MAIL SURVEY

A. IRB REVIEW

Prior to data collection, all materials (instrument, letters, fact sheets, etc.) were provided to the Westat IRB for review. In order to ensure informed consent, the IRB required the addition of several paragraphs to the survey cover describing the study and the use of the study data. On November 1, 2007, Westat's IRB approved the CTS survey data collection materials and procedures (see Appendix B, Attachment A). HSC also requested that the Westat IRB provide oversight for their work on the analysis and release of the database to a public website. On November 27, 2007, the Westat IRB agreed to serve as the approving body for HSC's work after the end of the official contract period (see Appendix B, Attachment B). HSC agreed to submit a checklist on Disclosure Potential of Proposed Data Releases as part of the IRB approval process.

Appendix B includes the final instrument (attachment C), cover letters for the first mailing (attachment D), cover letters for the second mailing (attachment E), cover letters for the third mailing (attachment F), fact sheets that accompanied the cover letters (attachment G), and a letter of support from the Agency for Healthcare Research and Quality (AHRQ, attachment H). The cover letters for the first mailing varied in the amount of incentive included as part of an experiment discussed in the next chapter ($50 versus $75), and the cover letters to the third mailing varied by incentive amount and by whether or not the physician had earlier cashed a check without responding.[5]

[5] Note that these documents all refer to the survey as the Community Tracking Study Physician Survey rather than the HSC Health Tracking Physician Survey as decisions regarding the name change had not as yet been made.

B. SAMPLE RELEASE

The preliminary results from the screening portion of the survey, conducted by Westat, were provided to MPR and an initial release was developed based on these results. The first release contained 9,478 physicians. This sample of 9,478 physicians included physicians who were screened as eligible for the mail survey as well as physicians who were determined by the screening as ineligible and physicians who had informed AMA not to include them in any survey. The working sample consisted of 8,367 physicians. The working sample excluded physicians who traced as ineligible or who were designated as "do not contact" by AMA, or who had been part of another sample.[6] Based on a preliminary assessment of expected response eligibility rates by strata and target numbers of completed interviews, we prepared a second sample release that included 772 physicians. The working sample released for data collection included 798 physicians (excluding 77 physicians who were classified as ineligible during the initial screening or who were designated as "do not contact" by AMA from among the 772 physicians and reintroducing the 103 physicians who were previously not released because they were deemed as potentially in the CATI sample). The final sample size for the mail survey was 10,250 (9,478 in the first release and 772 in the second sample release). The sample allocation is shown in Table V.1.

[6] We had originally planned on a telephone component for the 2008 survey to directly measure mode effects, but dropped this component after a pilot study indicated that the response rate from a telephone survey would be unacceptably low.

TABLE V.1

FINAL SAMPLE COUNTS FOR THE 2008 HEALTH TRACKING PHYSICIAN SURVEY

AMA Region	Final Sample Count		
	Total	PCP	Specialist
Total	10,250	4,271	5,979
1	732	287	445
2	945	357	588
3	834	349	485
4	1,133	468	665
5	1,282	475	807
6	1,225	497	728
7	886	399	487
8	864	359	505
9	1,095	496	599
10	1,254	584	670

Source: MPR computation.

C. EXPERIMENT TO TEST AMOUNT OF INCENTIVE AND FOLLOW-UP PROTOCOL

Because of the abandonment of the CATI portion of the survey due to low response rates, along with evidence of declining response rates on recent physician surveys conducted by HSC and other organizations, we conducted an embedded experiment to test the impact of differing levels of monetary incentive and follow-up efforts on response rates and survey costs. The experimental sample was "embedded" in the survey sample because the experimental cases comprised a significant part of the total sample. The results were then used to adopt an optimal incentive and follow-up protocol for the remainder of the survey.

A sample of 5,554 was used to test the amount of incentive ($50 vs. $75) and use of follow-up calls to respondents (received telephone calls vs. did not receive calls) on response and yield

rates. The response rate, which was the same as for the full study, was defined as completes plus ineligibles divided by the total sample, and the yield rate (which excludes ineligibles) was defined as complete eligible physicians divided by the total sample. The original design of the experiment was for 3,465 cases to be released in three arms: 1) $50 incentive with telephone follow-up, 2) $75 incentive with telephone follow-up, and 3) $75 incentive with no telephone follow-up. Because an early draft of the advance letter was erroneously sent to physicians in the last two experimental groups, and because this error had the potential of confounding the results of the experiment, two additional arms (2a and 3a) were added to the experiment. The size and original treatment of each group are shown in Table V.2 below. The experimental sample included cases that were deemed ineligible during screening and final refusals; these cases, which were not mailed questionnaires, were included in order to compute response rates based on the full sample. The experiment was conducted during the early stages of the field period, from February 5, 2008 through April 16, 2008.

TABLE V.2

EXPERIMENT SAMPLE BREAKDOWN

	Initial treatment	Sample size
Arm 1	$50, follow-up	2,054
Arm 2	$75, follow-up	702
Arm 3	$75, no follow-up	670
Arm 2a	$75, follow-up	1,074
Arm 3a	$75, no follow-up	1,054
Total Experiment		5,554

The experimental results were analyzed based on survey dispositions on May 2, 2008 (see Table V.3). Arms 2 and 3 were not used in the analysis because of the error in the advance letter, although the results were similar to arms 2a and 3a. A review of survey outcomes indicated that the mix of the $75 incentive and the follow-up calls (arm 2a) yielded significantly higher response and yield rates than the $75 incentive without follow-up or the $50 incentive with follow-up. An additional factor in assessing the three arms was the cost of the follow-up protocol, which was considered acceptable in order to achieve a higher response rate. The cost of the follow-up effort was modest as the protocol typically resulted in interviewers having brief calls with office staff or leaving messages, rather than speaking directly with physicians. Consequently, we decided to use the $75 incentive with follow-up calls for the remainder of the sample. A follow-up protocol was subsequently incorporated into additional mailings for arms 3 and 3a.

TABLE V.3

RESULTS OF EMBEDDED EXPERIMENT (MAY 2, 2008)

	Arm 1	Arm 2a	Arm 3a
Mailed Questionnaire	$50FU	$75FU	$75noFU
No Response	618	292	436
Survey Complete	866	498	438
Case Review - Interim[7]			
Re-mailed M1	60	27	1
Re-mailed M2	7	6	
Refused - Final	194	75	5
Deceased	2		
Ineligible - Invalid Specialty		1	
Ineligible - Federal Employee	8	2	5
Ineligible - Resident or Fellow	5		2
Ineligible - Less than 20 hours care	41	33	25
Ineligible, Other		1	
Ineligible - Not Practicing	3	1	
Ineligible - Retired	5	3	
Ineligible - Unavailable during Field period	2		
Address Unknown	84	46	36
Unable to Locate	45	27	33
Sub-Total	**1940**	**1012**	**981**
Final Disposition at Screening (No mailing)			
Scrn Inelig Fed Empl	4	3	3
Scrn Inelig Res/Fellow	7		1
Scrn Inelig No Direct Care	17	7	9
Scrn Inelig Specialty	2	1	
Scrn Inelig Retired	48	17	25
Scrn Inelig Not Avlb Fld Period	4	2	2
Scrn Inelig Institutionalized	4	11	3
Scrn Inelig Not in Prac	2	6	6
Scrn Inelig Other			
Scrn Deceased	4	3	1
Scrn Refusal	22	12	23
Sub-Total	**114**	**62**	**73**

[7] During follow-up calls, some respondents for physicians' offices requested that questionnaires be re-mailed. At the close of the experiment, 88 questionnaires had been re-mailed once and 13 had been re-mailed twice but had not yet been returned, to close out as final non-responses.

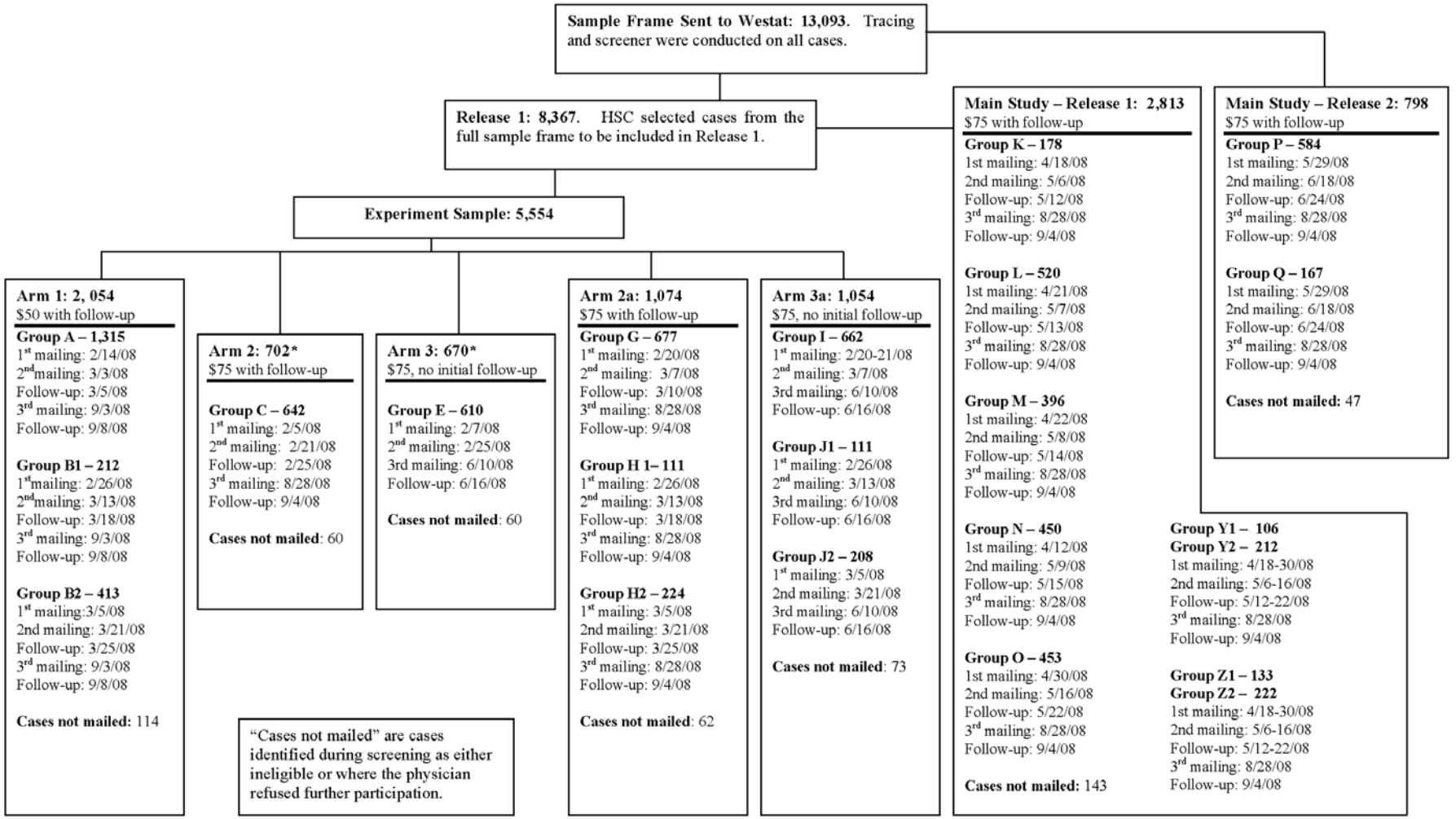
Total Sample	2054	1074	1054
Response Rate (C+I)/Total sample	49.85%	54.84%	49.34%
Yield Rate (C/Total sample)	42.16%	46.37%	41.56%

D. MAIL SURVEY DATA COLLECTION

The field period for the study started on February 5, 2008, and continued through October 31, 2008. This section outlines the procedures for the mailings, including package contents, calls from study respondents, and sending additional questionnaires by mail, fax and email. For exact dates of all mailings and follow-up calls see Exhibit V.1.

EXHIBIT V.1

STUDY DESIGN CHART



*These Arms originally included Groups Y and Z, which were moved to the Main Study-Release 1.

1. Mailing Protocol

Initial mailings to physicians were made in groups according to a schedule based on Westat corporate capabilities (i.e., check request requirements) and staff availability. Experimental groups received initial mailings in February 2008 and March 2008 while groups in the remainder of Release 1 received initial mailings in late April 2008. Groups in Release 2 received initial mailings in late May 2008. Second mailings were sent to physicians who had not responded within two weeks of the initial mailing. Physicians who sent back the first questionnaire did not have any further contact with the study. Physicians in the two experimental groups that did not initially receive follow-up treatment (arms 3 and 3a) received a third mailing that included follow-up calls in early June 2008. As a result of the success of this third mailing, the rest of the experimental groups (arms 1, 2 and 2a) and cases in the rest of Release 1 and Release 2 received a third mailing in September 2008. The number of cases to receive specific mailings is outlined in Table V.4 below.

TABLE V.4

NUMBER OF CASES PER MAILING

		First mailing		Second mailing	Follow-up call to second mailing	Third mailing		Follow-up call to third mailing
		$50 check	$75 check			No check	$75 check	
Experiment	Arm 1	1,940	0	1,328	Yes	59	347	Yes
	Arm 2	0	642	421	Yes	54	297	Yes
	Arm 2a	0	1,012	682	Yes			Yes
	Arm 3	0	610	397	No	12	682	Yes
	Arm 3a	0	981	661	No			Yes
Release 1		0	2,670	1,839	Yes	77	679	Yes
Release 2		0	751	489	Yes	30	170	Yes
Total		1,940	6,666	5,817		232	2,175	

2. Mailing Methods

For the experimental groups, two mailing methods were used. Cases that were coded as complete at the screener level were sent packages using two-day Federal Express delivery. Cases that were not coded as complete at the screener were sent packages using the US Postal Service's (USPS) Priority Mail system, which takes four days for delivery. USPS Priority Mail was used for screener non-completes because this delivery method obtained address corrections if the screener address was incorrect. Due to escalating costs for Federal Express delivery, all remaining cases in Release 1 and Release 2, as well as all third mailings, were sent using Priority Mail. Beginning April 18, 2008, all re-mails (requests from physicians to receive an additional questionnaire package) were also made by Priority Mail.

3. Package Contents

Table V.5 below summarizes the contents of each mailing. Initial mailings to respondents included:

- a questionnaire labeled with respondent and form IDs;
- a personalized cover letter (on Robert Wood Johnson Foundation letterhead);
- a fact sheet;
- a letter of support from the Agency for Healthcare Research and Quality;
- an incentive check of either $50 or $75 depending on group assignment; and
- a postage-paid return envelope.

TABLE V.5

CONTENT OF MAILINGS

	First mailing	Second mailing	Third mailing		Re-mails	
			Non-cashers	Cashers	Full package	Check only
Questionnaire	✓	✓	✓	✓	✓	
Cover Letter	✓	✓	✓	✓	✓	✓
Fact Sheet	✓	✓	✓		✓	
Letter of Support	✓	✓	✓		✓	
Incentive Check	✓		✓			✓
Return Envelope	✓	✓	✓	✓	✓	

Each package was assembled with the return envelope folded in half and inserted into the front cover of the questionnaire with the incentive check, cover letter, letter of support and fact sheet cradled inside the fold of the return envelope. This arrangement was designed to ensure that incentive checks did not get separated from the rest of the materials. Second mailings and re-mails included all of the materials in the first mailing with the exception of the incentive check; the cover letter for the second mailing also was altered to reflect the absence of the check. Third mailings were divided between non-respondent physicians who had not cashed their incentive check and those that had. Physicians who had not cashed their check were sent a completely new package in the third mailing, including a new incentive check. Physicians in this group that had previously received a $50 incentive check were sent a $75 check. Physicians who had cashed their earlier incentive check but had not responded to the survey were sent a new

package without a new check. These cases also received a different cover letter. All versions of cover letters, fact sheets, and letters of support can be found in Appendix B.

4. Undeliverables

Two types of packages were returned to Westat as undeliverable: those with and those without a forwarding address. If new or forwarding address information was provided by the mailing service, this new address information was entered into the SMS and the package was resent to the new address. If no forwarding address was provided, the case was marked in the SMS as "address unknown" and sent to the TRC for tracing. If no new address could be located during telephone tracing, the case was finalized as "non-locatable." A total of 917 cases were finalized as non-locatable.

5. In-Bound Calls

Materials sent to respondents referenced a toll-free number that respondents could call for additional information. At the beginning of the field period, calls went directly to The Robert Wood Johnson Foundation (RWJF). However, it became clear that most calls could be more easily answered by a staff member from Westat. Therefore, a toll-free number connecting to Westat was established. However, if the nature of the call indicated that the physician wanted to discuss the study rationale or had questions about the use of the data, the call was referred to the RWJF.

A total of 113 calls were received on the Westat toll-free line during the course of the study. Each call was logged into a tracking system and appropriately handled by a study staff person. The reasons for the calls and the protocol response are outlined in Table V.6 below. It should be

noted that some calls covered more than one topic and have therefore been included in the table more than once. This table only includes calls received directly by Westat; it does not include calls made to the RWJF.

TABLE V.6

DESCRIPTION OF INCOMING CALLS TO THE WESTAT TOLL-FREE NUMBER

Number of Calls	Reason for call	Protocol response
30	Physician cannot find honorarium and requests that a new check be sent	A new check was mailed to the physician.
16	Report that the physician is not eligible for the study.	The physician was asked to complete the screener portion of the instrument and send it in.
7	Refusal to participate; request no additional mailings or calls.	The case was marked as a hard refusal and received no additional mailings or calls.
10	Physician has moved.	If a new address was provided, a new package was mailed. If no new address was provided, the case was marked as "non-locatable."
7	Questionnaire already sent in, but received another in the mail and wants to know whether to fill it out again.	Physicians were assured that they did not need to complete the survey a second time.
11	Want to confirm that the mailed questionnaire has been received.	The case was checked in the SMS and receipt confirmed.
34	Other question or request.	Response varied according to need.

6. Re-mails

A report was run daily to indicate which cases required a re-mail, fax or email. The majority of these requests originated in the TRC, although they could also be initiated by either the in-

bound toll-free number or through an undeliverable package with a corrected address. Re-mail requests were assigned a status in the management system that indicated whether the physician needed a repeat of either the first or second package. Physicians requesting the first package were sent the full initial package including a new check. Physicians requesting the second mailing received a complete data collection package except for a new check. An additional type of re-mail was for a check only. Physicians who reported that they had misplaced their incentive check were sent a new check. Table V.7 below outlines the number of re-mails by type in each of the sample groups.

TABLE V.7

TYPES OF RE-MAILS

		First package re-mails	Second package re-mails	Check-only re-mails
Release 1 - Experiment	Arm 1	144	27	24
	Arm 2	38	15	8
	Arm 3	33	3	5
	Arm 2a	77	15	9
	Arm 3a	68	4	5
Release 1 - Remainder		219	20	31
Release 2		84	1	6
Total		663	85	88

7. Fax and Email Requests

Physicians could request to have the survey either faxed or emailed to them. These requests came through either the TRC or the in-bound toll-free number. Physicians requesting a fax received a set of materials including a questionnaire with ID number, a cover letter, and a customized fax cover sheet. They were asked to either mail the questionnaire back to the

provided address or fax it back to the provided number. Physicians requesting an email received the same set of materials in electronic (PDF) format. Physicians were asked to print the questionnaire before completing it and either mail it or fax it back. The number of fax and email requests is outlined in Table V.8. Some physician offices requested both a fax and an email or requested a fax on more than one occasion; therefore, Table V.6 reports the number of requests rather than the number of cases.

TABLE V.8

NUMBER OF FAX AND EMAIL REQUESTS

		Fax requests	Email requests
Release 1 - Experiment	Arm 1	99	30
	Arm 2	15	1
	Arm 3	21	3
	Arm 2a	67	18
	Arm 3a	80	0
Release 1 - Remainder		219	13
Release 2		62	5
Total		563	70

8. Mail Refusals

Packets returned with blank questionnaires and the original mailing contents (with or without the check) were deemed final refusals and were marked in the SMS as “final refusal” along with other comments that indicated the doctor was unwilling to complete the survey.

E. TELEPHONE FOLLOW-UP

Except for experimental arms 3 and 3a, follow-up telephone calls were made to the physician's office following the second and third mailings in order to encourage survey completion. (Based on analysis of experimental results, arms 3 and 3a received follow-up calls after the third mailing). Calls were made to physicians' offices between February 25 and October 31, 2008. The protocol for the follow-up calls is outlined in this section.

1. Training Interviewers

Fifteen interviewers, selected based on prior experience with physician studies that used similar methods, were trained on February 25, 2008. The training included an overview of the web-based Study Management System (SMS) used for tracking cases in the TRC. During the training, a total of six interactive practice sessions were conducted and a debriefing was held after the first hour of production to address any questions or issues.

2. Telephone Follow-up Protocol for Cases that were Complete at Screening

Cases coded as screener completes that did not respond to the initial mailings were assigned to the full telephone follow-up protocol outlined in this section. If a completed questionnaire was received at any time during the follow-up period, calls were halted and the case was coded as complete. Calls were scheduled as follows:

If the last mailing date was…..	Then follow-up began the following:
Monday	Friday
Tuesday	Monday
Wednesday	Tuesday
Thursday	Wednesday
Friday	Thursday

Interviewers were instructed to avoid calling offices between the hours of noon and 1:00 pm in the respondent's time zone to avoid the lunch hour. Later, it was discovered that these times varied from office to office. Therefore, those rules were modified according to notes provided for those offices that were considered outside the norm.

The SMS used to track cases in the TRC provided interviewers with a special section to note any unique instructions related to that particular office or case. These notes were routinely reviewed by the TRC manager to ensure that any particular or irregular circumstance for each case was addressed appropriately.

Follow-up calls were broken down into 3 tasks:

- **Task 1:** The goal of Task 1 was to verify that the package was received at the physician's office and to ask that it be physically handed to the doctor. In the case of small practices, this call was used to remind the physician to fill out the questionnaire. Up to five calls were made in trying to complete Task 1. If Task 1 was not completed in five calls, the TRC supervisor determined whether to move the case to Task 2 or to continue trying to complete Task 1.

- **Task 2:** Calls to complete Task 2 were made 7-10 days after the completion of Task 1. The goal of Task 2 was to leave a reminder message for the physician and to ascertain whether an additional survey needed to be sent. Two calls were made in an attempt to

complete Task 2. If Task 2 was not complete in two calls, the TRC supervisor determined whether to move the case to Task 3 or to continue trying to complete Task 2.

- **Task 3:** Task 3 calls started 10 days after the previous call. The goal of Task 3 was to speak with the physician directly. If this was not possible, Task 3 enlisted the help of the administrative assistant or office manager to encourage the physician to complete the survey. The interviewer also determined whether an additional survey needed to be mailed or faxed. A total of three calls were made in an attempt to complete Task 3. If Task 3 was not complete in three calls, the TRC manager reviewed the case to decide if additional calls would be helpful or if the case had been completely worked.

No more calls were made to the physician's office once the case was finalized or Task 3 was completed.

3. Telephone Follow-up Protocol for Cases other than Complete at Screening

Cases that were coded as "non-locatable" during the screener were not sent to the TRC for follow-up calls. Cases that had screener codes of maximum calls, no contact, or refusal were sent to the TRC for a single follow-up call. If this call was successful (e.g., reached a human being), then that case entered the full telephone follow-up protocol. If the single call was not productive, no further follow-up was attempted.

4. Telephone Follow-up to Re-mail, Fax, and Email Requests

Physicians who received re-mails were returned to the TRC for follow-up once the additional materials were sent. These physicians received a follow-up call the fourth day after the package was re-mailed. Respondents requesting a check received a follow-up call seven days after the check was mailed. No follow-up calls were made if the completed questionnaire was received before the follow-up was scheduled to begin.

5. Calls Following the Third Mailing

Following the third mailing, a limited follow-up telephone protocol was used to encourage survey participation and to verify the mailing had been received. Each case was called only enough times to either talk to a person to confirm package receipt or to leave a message on an answering machine. Cases received no more than four total follow-up call attempts after the third mailing. The only exceptions were cases where a person who was reached requested that the package be re-mailed (either by Priority Mail, fax or email). Each of these cases received another follow-up call to confirm that the additional package was received.

6. Refusals by Telephone

Physicians who spoke to data collectors or called the toll-free number and stated they did not want to complete the survey were coded as final refusals and no additional follow-up calls were made to them. If someone other than the sampled physician called on behalf of the physician and stated that the physician did not want to complete the survey, the call history was reviewed by a supervisor. If the supervisor determined that a return call would be unproductive, the case was entered as a final refusal. Otherwise, the case received additional efforts according to the telephone follow-up protocol.

F. DATA MANAGEMENT AND PROCESSING

1. Receipt of Questionnaires

Completed questionnaires were processed and entered into the SMS on the same day or the next morning after arrival. Because each respondent could receive more than one mailing,

questionnaires were entered into the system by both the respondent ID and the form ID. This ensured that any duplicate questionnaires from the same respondent were noted and one could be discarded during cleaning. Questionnaires returned as undeliverable were also entered into the SMS as "Address Unknown" for the first two returned-as-undeliverable mailings. Upon receipt of the third returned mailing, the case status was changed to a final status of "Unable to Locate."

All returned questionnaires were grouped and filed according to a status code (complete, ineligible, address unknown, etc). Completed questionnaires were coded, verified, batched and sent to data entry.

2. Coding and Keying

Completed questionnaires were coded according to decisions made jointly by HSC and Westat staff. Comments written in the margins were flagged and then reviewed by a supervisor to determine relevance to questionnaire responses. Problematic responses were also flagged and reviewed by a supervisor. All decisions regarding how to code responses were documented in a decision log and resultant new codes and consistency checks were incorporated into the codebook.

The supervisor reviewed the first 25 questionnaires completed by each coder. If there were no coding issues identified, the supervisor reviewed 10 percent of all subsequent coding work.

Following coding, questionnaires were batched and sent to data entry where they were double keyed. Once questionnaires were returned from data entry, iterative editing resumed, during which frequencies and cross tabs were generated and reviewed regularly. Comments written by physicians concerning the survey or health issues that they would like to see

addressed in future surveys were also captured in a separate Excel file. Completed questionnaire data was reconciled with the disposition codes recorded in the management system.

In accordance with IRB requirements, identifying information located on the last page of the questionnaire was keyed into a file kept separate from the rest of the data.

3. File Layout and Development

A SAS file of cleaned questionnaire data was created for delivery to HSC. This file contained SAS special missing values to indicate "Don't know" (D), "Not Ascertained" (M), and "Not Applicable" (.) values.

4. Data Delivery

An interim data file was delivered to HSC in August 2008 and final data files were delivered in November 2008. The interim delivery consisted of 3,637 eligible records from the physician mail survey. The final data file produced by Westat consisted of 4,723 eligible physician records; subsequent editing reduced this number to 4,720. Both data deliveries consisted of SAS records and were accompanied by the following: files for format linking, formats, and frequencies; a status file related to the screening, mailing and follow up processes; an Excel file of physicians statements from the comment section at the end of the questionnaire; an Excel file containing the name and address for each physician's main medical practice and the hospital name where he or she admits the most patients; and a codebook of the mail questionnaire.

5. Final Dispositions and Response Rate

The final dispositions of the sample and response rate are shown in Table V.9. The unweighted response rate, which is defined as the sum of completed eligible and ineligible cases divided by the total sample size, is 61.8 percent; the weighted response rate, shown in Table VI.1 of Chapter VI below, is 61.9 percent. This definition of the response rate assumes that the eligibility rate for non-responding and non-locatable physicians is the same as for responding physicians. Since the eligibility rate for non-locatable physicians is likely to be lower than for locatable physicians, this definition is conservative.

TABLE V.9

FINAL DISPOSTIONS AND RESPONSE RATE

Mail Survey Final Dispositions	**Total Sample**
No Response	1,460
Survey Complete	4,720
Refused - Final	1,104
Ineligible-Deceased	10
Ineligible - Invalid Specialty	4
Ineligible - Federal Employee	45
Ineligible - Resident or Fellow	17
Ineligible - Less than 20 hours care	257
Ineligible - Not Practicing	36
Ineligible - Retired	29
Ineligible - Unavailable during Field period	7
Ineligible, Other	0
Address Unknown	0
Unable to Locate	917
Total Mail Release	**8,606**
Screening and Tracing Final Dispositions	
Ineligible- Fed Employee	98
Ineligible- Resident/Fellow	138
Ineligible-Less than 20 hours care	238
Ineligible- Specialty	24
Ineligible- Retired	400

Ineligible- Not Avlb Fld Period	51
Ineligible- Institutionalized	81
Ineligible- Not in Practice	107
Ineligible- Other	1
Final Refusal	132
Ineligible-Deceased	59
Ineligible (from CATI sample)	9
Screening Final	**1,338**
AMA Refusals (not attempted)	**306**
Total	**10,250**
Complete	**4,720**
Ineligible	**1,611**
Refusal (Eligibility Unknown)	**1,236**
No Response (Eligibility Unknown)	**2,683**
Total Sample	**10,250**
% Eligible	**74.55%**
Eligible / (Eligible + Ineligible)	
Estimated Eligibles	**2,922**
Total Eligibles	**7,642**
AAPOR Response Rate (equivalent to (completes+ineligibles)/total sample	**61.8%**

VI. SAMPLING AND ANALYSIS WEIGHTS

A. OVERVIEW

The weights for the 2008 Health Tracking Physician Survey adjust for differences in probabilities of selection and response (that is, the propensity for a physician to be located and the propensity for a located physician to respond). The initial weights, also called sampling weights, were calculated as the reciprocals of the probabilities of selection. The initial weights were adjusted to account for locatability and non-response because some sampled physicians could not be located and others that were located did not participate.[8] After these non-response adjustments, the weights were post-stratified. In this section we describe the initial weights, non-response patterns that motivate adjustments, and the adjustments themselves.

1. Analysis Weights

Unbiased estimates are the goal of any serious survey. Differences in probabilities of selection or response propensities across various population subgroups can result in the responding sample being distributed differently than the study population. Such inconsistent distributions, if not corrected by proper weighting, can produce biased survey estimates. Thus, our analysis weights adjust for differences in selection probabilities and the two components of response: 1) locatability and 2) participation among physicians that could be located. To calculate the adjustments for locatability and non-response, we employed logistic regression

[8] For the purposes of both the examination of non-response and the weighting adjustments, "participation" and "response" include those determined in the course of the survey to be ineligible (such as those that had retired or whose practice included fewer than 20 hours per week of patient contact), as well as those that completed the questionnaire.

models using data from the AMA Masterfile (the sampling frame). Separate models were developed for each of the two adjustments.

B. INITIAL WEIGHTS

The initial sampling weight was calculated as the reciprocal of the probability of selection of each physician. The sample was selected and released for contact as described in Chapter III, and the probabilities of selection reflected each of the steps in selecting and releasing the sample. Probabilities of selection varied only slightly across strata. Thus, the sampling weights were roughly equal for all sample members.

C. RESPONSE PATTERNS

Response patterns were examined to assess the potential for non-response bias, gauge whether adjustment should be made in one or more steps, and inform the selection of variables to include in non-response adjustment models. We used data available from the AMA Masterfile to evaluate the response patterns. First, we concluded that response rates differed across groups of sampled physicians defined by characteristics that could be related to study variables; these differences, if not incorporated into weighting adjustments, could produce biased survey estimates. Second, patterns for the two components of non-response, locatability and propensity to respond once located, were sufficiently different to warrant separate adjustments.

On a weighted basis, 61.9 percent of the sample responded (i.e., completed the survey or were determined to be ineligible). The (net) percent responding is the product of the weighted percent located (88.1) and the weighted percent of those located who responded (70.3). We

found variation in all three measures across subgroups of the population that were defined based on frame information. We examined response patterns for subgroups based on:

- Classification (whether PCP or specialist)
- Region
- Gender
- Age
- Country of birth (United States or Canada; all others)
- Medical school location (United States or Canada; all others)
- Specialty among the PCPs (General/Family Practice, Internal Medicine, Pediatrics)
- Practice arrangement (solo- or two-physician practice; office, group, or HMO; all others)
- Percent of time practicing in hospitals
- Survey incentive

We examine each of these in turn. Results are presented in Table VI.1.

Classification and Region. A slightly higher net percentage of PCPs responded (62.6) than did specialists (61.4), but the patterns differed and there was more variation in the two components than in the net response. PCPs had a lower percentage located (86.7 versus 89.1) but a higher percentage responding among those located (72.2 versus 69.0). Across regions, the net percent responding fell between 57.9 and 67.4. The percent located ranged from 84.9 to 90.6 while the percent responding among those located ranged from 66.7 to 76.3.

When classification and region are crossed, they define the sampling strata. We find somewhat larger differences across strata, with the net percent responding ranging from 56.2 to 69.2, the percent located from 82.6 to 92.0, and the percent responding among those located from 63.7 to 78.1.

TABLE VI.1

RESPONSE PATTERNS FOR THE 2008 HEALTH TRACKING PHYSICIAN SURVEY

Sample Classification	Total Sample	Unweighted Located Sample	Weighted Percent Located	Unweighted Sample Completes and Ineligible	Unweighted Sample Completes	Weighted Response Among Located	Weighted Percent Response
TOTAL	**10,250**	**9,027**	**88.1**	**6,331**	**4,720**	**70.3**	**61.9**
Physician Classification							
PCP	4,271	3,699	86.7	2,665	1,959	72.2	62.6
Specialist	5,979	5,328	89.1	3,666	2,761	69.0	61.4
Region							
1	732	630	86.1	434	309	68.9	59.3
2	945	802	84.9	547	410	68.2	57.9
3	834	729	87.5	526	395	72.1	63.1
4	1,133	1,004	88.7	714	518	71.1	63.0
5	1,282	1,160	90.3	816	605	70.4	63.6
6	1,225	1,110	90.6	741	560	66.7	60.5
7	886	799	90.2	557	420	69.6	62.8
8	864	762	88.2	514	403	67.7	59.7
9	1,095	968	88.4	739	565	76.3	67.4
10	1,254	1,063	84.9	743	535	70.0	59.4
Sampling Strata (Classification and Region)							
101	287	246	85.7	179	135	72.8	62.4
102	357	295	82.6	210	153	71.3	58.8
103	349	300	86.0	221	164	73.7	63.3

Sample Classification	Total Sample	Unweighted Located Sample	Weighted Percent Located	Unweighted Sample Completes and Ineligible	Unweighted Sample Completes	Weighted Response Among Located	Weighted Percent Response
104	468	403	86.1	294	207	73.0	62.8
105	475	418	88.0	301	217	72.0	63.4
106	497	446	89.7	290	215	65.0	58.4
107	399	351	88.0	261	190	74.4	65.4
108	359	316	88.0	230	188	72.8	64.1
109	496	439	88.5	343	259	78.1	69.1
110	584	485	83.0	336	231	69.3	57.5
201	445	384	86.3	255	174	66.4	57.3
202	588	507	86.2	337	257	66.5	57.3
203	485	429	88.5	305	231	71.1	62.9
204	665	601	90.4	420	311	69.9	63.2
205	807	742	91.9	515	388	69.4	63.8
206	728	664	91.2	451	345	67.9	62.0
207	487	448	92.0	296	230	66.1	60.8
208	505	446	88.3	284	215	63.7	56.2
209	599	529	88.3	396	306	74.9	66.1
210	670	578	86.3	407	304	70.4	60.7

Sample Classification	Total Sample	Unweighted Located Sample	Weighted Percent Located	Unweighted Sample Completes and Ineligible	Unweighted Sample Completes	Weighted Response Among Located	Weighted Percent Response
Gender							
Male	7,363	6,585	89.5	4,618	3,470	70.3	62.9
Female	2,887	2,442	84.6	1,713	1,250	70.4	59.6
Age							
20-44 years	3,301	2,841	86.1	1,994	1,600	70.3	60.5
45-54 years	3,301	2,945	89.2	1,957	1,591	66.6	59.3
55-64 years	2,359	2,097	88.9	1,443	1,113	69.1	61.5
65 years or older	1,289	1,144	88.9	937	416	82.1	73.0
Gender and Age							
Male, 20-44 years	1,932	1,683	87.1	1,191	987	70.9	61.8
Male, 45-54 years	2,334	2,127	91.2	1,388	1,172	65.3	59.5
Male, 55-64 years	1,944	1,741	89.6	1,197	931	69.0	61.8
Male, 65 years or older	1,153	1,034	89.8	842	380	81.6	73.3
Female, 20-44 years	1,369	1,158	84.7	803	613	69.5	58.8
Female, 45-54 years	967	818	84.3	569	419	69.9	59.0
Female, 55-64 years	415	356	85.8	246	182	69.8	59.9
Female, 65 years or older	136	110	81.6	95	36	86.4	70.5
Birth Country							
U.S.	7,190	6,397	89.0	4,580	3,448	71.8	63.9
Other	3,060	2,630	86.0	1,751	1,272	66.7	57.4
Medical School Location							
U.S./Canada	7,882	6,987	88.6	4,966	3,740	71.2	63.2

Sample Classification	Total Sample	Unweighted Located Sample	Weighted Percent Located	Unweighted Sample Completes and Ineligible	Unweighted Sample Completes	Weighted Response Among Located	Weighted Percent Response
Other	2,368	2,040	86.2	1,365	980	67.1	57.8
Specialty							
General/family practice	1,715	1,495	87.2	1,073	790	71.9	62.7
Internal medicine	1,624	1,395	86.0	939	679	67.5	58.1
Pediatrics	932	809	86.9	653	490	80.9	70.3
Specialist	5,979	5,328	89.1	3,666	2,761	69.0	61.4
Present Employment							
Solo or 2 practice	2,487	2,275	91.5	1,607	1,171	70.8	64.8
Office-group-HMO'	4,389	3,985	90.8	2,778	2,242	69.9	63.4
Other	3,374	2,767	82.0	1,946	1,307	70.6	57.9

Sample Classification	Total Sample	Unweighted Located Sample	Weighted Percent Located	Unweighted Sample Completes and Ineligible	Unweighted Sample Completes	Weighted Response Among Located	Weighted Percent Response
Percent of Practice Hours Spent at Hospital							
0 percent or unknown	6,044	5,181	85.7	3,546	2,541	68.6	58.8
1-20 percent	1,075	1,017	94.6	789	656	77.7	73.6
25-45 percent	1,036	952	92.0	668	522	70.5	64.9
50-85 percent	1,001	910	91.0	618	479	68.2	62.0
86-100 percent	1,094	967	88.4	710	522	73.5	64.9
Incentive Experiment [9]							
$50, full protocol	2,325	2,030	87.3	1,385	1,031	68.4	59.8
$75, full protocol, old letter	1,164	1,028	88.3	730	535	71.2	62.9
$75, no follow-up, old letter	1,164	1,010	86.7	714	543	70.9	61.4
$75, full protocol, new letter	1,214	1,071	88.2	779	585	73.1	64.5
$75, no follow-up, new letter	1,201	1,081	90.1	759	574	70.3	63.3
Not in the experiment	3,182	2,807	88.2	1,964	1,452	70.1	61.9

Source: MPR computations

[9] Note that the experimental protocols were standardized following analysis of the experimental data (discussed above in Chapter V). The $75 protocols with no follow-up received follow-up calls for the third mailing and physicians who had not responded to the $50 protocol after the second mailing were mailed $75 for the third mailing. Physicians who were not part of the experiment all received the $75 follow-up protocol. The incorrect (old) letter was replaced with a corrected letter for the second and third mailings.

Age and Gender. The net percent responding was higher for male (62.9) than for female physicians (59.6). Male physicians were also more likely to be located (89.5 versus 84.6 percent), possibly because of name changes, while the percent responding among those located was nearly identical for the two groups.

Response rates varied by age, with the percent responding being substantially higher for those 65 and older (73.0) compared with younger cohorts (59.3 to 61.5 percent). The percent located did not vary greatly, but was lower for the youngest group (86.1 compared to approximately 89 for the other three). There was more variation in percent responding among those located, which ranged from 66.6 (age 45-54) to 82.1 (age 65+).

When age and gender are crossed, the net response rates range from 58.8 percent (female physicians age 20-44) to 73.3 percent (male physicians, age 65+). Compared to female physicians, the net response for males is two to three percentage points higher in all age groups except 45 to 54 year olds. The percent located varies from 81.6 to 91.2, while the percent responding among those located has an even larger range (65.3 to 86.4).

Birth Country and Medical School Location. Compared to physicians born outside the U.S., overall response was higher for native-born physicians (63.9 versus 57.4 percent). Similarly, the net response was higher for those who attended medical school in the U.S. or Canada than for others (63.2 percent compared to 57.8 percent). In both cases the gap was larger for response among those located than it was for the percent located.

Medical Practices. In addition to distinguishing broadly between PCPs and specialists, we examined response patterns across three specialties within primary care, practice arrangement, and percentage of hours spent practicing in a hospital setting, and noted differences in response patterns along all dimensions.

While a slightly higher percentage of PCPs than specialists responded, there was considerable variation among the three subgroups of PCPs. The net percent responding ranged from 58.1 (internal medicine) to 62.7 (general and family practice) to 70.3 (pediatrics). There was little variation among groups with respect to percent located, but a large spread (13.4 percentage points) in the percent responding among those located.

We find a smaller difference in net percent responding by practice arrangement than across specialties. Net rates vary from 57.9 (other) to 64.8 percent (solo practice/two-physician practice). Unlike the difference across specialties, the difference across practice type is mostly attributable to locatability, where there is a 9.5 percentage point spread, compared to response, where there are almost no differences among the three types.

Examination of percent of hours practicing at a hospital shows a spread of roughly 15 percentage points in net response. This variable was derived from another survey conducted by the AMA, results from which were included on the Masterfile. The highest net response corresponded to those spending 1-20 percent of their time in a hospital, and the lowest to those with zero percent or missing data. The 1-20 percent group had the highest percent located and highest percent responding among those located. The zero (or unknown) percent group had the lowest percent located and was nearly the lowest in percent responding once located.[10] Since the Masterfile did not differentiate between those values, it is possible that many of these cases did not respond to the AMA survey from which these data were derived, which could explain their lower response rates for the HSC survey.

[10] Fifty nine percent of the sample had a value of zero or unknown hours.

Survey Incentive. The variation in net response among incentive groups was less than that for groups defined on other criteria, ranging from 59.8 to 64.5 percent.[11] The spread was similar for percent responding once located, and a bit smaller (86.7 to 90.1) for percent located.

The differences in net response led us to conclude that using frame variables including those discussed above was appropriate when developing a model or models to adjust for non-response. Since different patterns were observed for percent located and percent responding once located, we developed separate models for locatability and for response propensity among those located.

D. WEIGHT ADJUSTMENTS

The purpose of non-response adjustment to sampling weights is to reduce the potential for bias associated with non-response. If non-response to a survey is completely random, then estimates of means weighted by the sampling weights would be unbiased and no adjustment would be necessary. For estimating totals, however, a single adjustment still would be needed to inflate a weighted total to account for the proportion of physicians who did not respond.

However, non-responses are rarely completely random, and examination of response patterns suggests it was not for the CTS Physician Survey. Our approach to non-response adjustments (consistent with the patterns noted above) was to develop two logistic regression models designed to predict (1) the likelihood of locating a physician (location propensity score) and (2) the likelihood that located physicians complete the interview (response propensity score). Then, we computed an adjustment value for each physician who completed the interview. The

[11] Note that the differences in experimental groups understate the effect of the higher incentive and follow-up interventions because all non-responding physicians who were selected for the experiment were given the preferred approach ($75 incentive with follow-up) in subsequent mailings.

weight as adjusted for non-response is the product of the inverse of the location propensity score, the inverse of the response propensity score, and the sampling weight.

A key factor in determining the usefulness of logistic regression models is the availability of information for respondents and non-respondents. In many surveys, information is limited beyond that used for creating sampling strata. However, 2008 Health Tracking Physician Survey has information for nearly all sampled physicians that can be used to enrich the models; the AMA file that was used as the sample frame contains many demographic and practice characteristics for physicians.

Logistic propensity modeling has been used for several surveys where information on the characteristics of both respondents and non-respondents is available. For example, this approach was used for the National Survey of Family Growth (Potter et al., 1998), and has been tested for use with the Survey of Income and Program Participation (Folsom and Witt, 1994). The procedure also has been employed in surveys of military personnel (Iannacchione et al., 1991) and in surveys of Medicare and Medicaid populations for which demographic and economic data are available from federal or state administrative files (CyBulski et al., 1999).

The modeling approach can result in a few sample members being assigned an extremely large adjustment factor (Little, 1986). However, the possibility of large adjustment factors can be reduced by using a restricted logistic regression model[12] or by trimming to compensate for adjustment factors from an unrestricted logistic regression model via a sample alignment or a post-stratification adjustment process. We used the latter approach. As discussed below, we examined the weights for outliers and concluded that trimming was not needed.

[12]The coefficients of the model are estimated based on restrictions on the size of the adjustment factor.

The model-based non-response adjustments are predicted values (based on maximum likelihood) and are estimators that are consistent, asymptotically efficient, asymptotically normal, and therefore, asymptotically unbiased.

After computing adjustment factors for the inability to locate a physician and for non-response among located physicians, these non-response adjusted weights were then checked for consistency with known (or estimated) population counts of eligible physicians and were post-stratified.

We prepared two sets of weighted logistic regression models to adjust the survey weights for our ability to locate physicians and to obtain a response (either a completed survey or ineligible disposition) among the located cases. Each model was used to predict locatability or response among located cases as a function of physician characteristics, represented by a series of indicator variables. The sampling weights were used in the location regression models and the sampling weights adjusted for non-location were used in the response regression models.

The variables used in the logistic regression models, chosen based upon the abovementioned non-response pattern analysis, were age, country of medical school, country of birth, gender, specialty, present employment, percentage of hours the physician worked in a hospital, year licensed, type of incentive offered, AMA region, and whether located in an MSA. We began by including all of them in the models (referred to as the full model). Many of these variables were in the form of multi-level categorical responses, so we transformed them into a series of indicator variables. To identify interaction terms among the main effects variables that should be included in the model, we employed the method of Chi-Square Automatic Interaction Detection (CHAID). Second-, third-, and fourth-order interactions were included if indicated by the CHAID analysis.

Nested models were used so that all lower-order interactions within a significant higher order interaction were included in the model regardless of their significance.

The categories for the first-order variables were chosen based on the number of observations in each category and the different location or response rates in each one. For example, the categorization of specialty in the model takes four categories: General/Family Practice, Internal Medicine, Pediatrics, and Specialists.

To prepare the models, we used a weighted, forward stepwise variable selection logistic regression procedure from SAS, which identifies and adds the predictor that minimizes the deviance when a new predictor is introduced in the model. We obtained a full logistic regression model with this method. Then we used this full model in SUDAAN, which computes accurate variances for the estimates of the models, taking into account the sampling design of the survey, and eliminated predictors that were not significant.

Table VI.2 summarizes the logistic regression model that was used for the location adjustments and Table VI.3 presents results from the response model. For each model, we also present the pseudo R-squared values, noting that small pseudo R-squared values are the norm in logistic regression and cannot be interpreted in the same way as those from linear regression (Hosmer and Lemeshow, 2000).

The goodness-of-fit tests indicated that the models were a reasonable fit. The pseudo R-squared values were small for some models, with an average value of 0.06 for the location model and 0.04 for the response model.

1. Location Weight Adjustments

The location models estimate the probability of locating a physician (location propensity score). The weight adjusted for location is obtained by multiplying the sampling weight and the

inverse of the location propensity score. These adjustments inflate the weights of the located physicians to compensate for those physicians who were not located.

The final logistic regression model for location showed that the location rates were higher among the following categories of physicians:

- specialists;
- those in the experimental arm receiving a survey incentive of $50 with the full follow-up protocol;
- those who did not respond/responded zero or who responded "85% or more" to an AMA survey question on the percent of hours spent in hospital care (odds ratio 6.5);
- those whose present employment was not in the "other" category (odds ratio 3.2); and
- those born in the USA or Canada and whose present employment was in the "other" category (odds ratio 1.4).

TABLE VI.2

RESULTS OF THE LOCATION MODELING PROCEDURES

Variables Included in the Final Model[a]
Medical school location (in USA or Canada; other)
Specialty (General, Family Practice, Internal Medicine, Pediatrics, Specialist)
Percent hours spent in hospital
Age (20-44, 45-54, 55-64, 65+)
Survey incentive
Physician's gender
Employment (solo or two-physician; larger practice, group or HMO; other)
Birth country (USA, Canada, or other)
Whether practice is in MSA (population $\geq$ 250,000) or not

[a]All variables were significant at $p \leq 0.3$ as main effects or within interactions.

TABLE VI.3

RESULTS OF THE RESPONSE MODELING PROCEDURES

Variables Included in the Final Model[a]
Medical school location (in USA, Canada; other)
Specialty (General, Family Practice, Internal Medicine, Pediatrics, Specialist)
Percent hours spent in hospital
Age (20-44, 45-54, 55-64, 65+)
Survey incentive
Physician's gender
Employment (solo or two-physician; larger practice, group or HMO; other)
Birth country (USA, Canada, or other)
Whether practice is in MSA (population $\geq$ 250,000) or not
Year of first license
Region

[a]All variables were significant at $p \leq 0.3$ as main effects or within interactions.

2. Response Weight Adjustments

The response models predict the probability that a physician completes the interview (response propensity score). The final weight adjusted for non-response is obtained by multiplying the weight adjusted for location and the inverse of the response propensity score. These adjustments inflate the weights of the physicians who completed the interview to compensate for those physicians who did not complete the interview.

The final logistic regression model for response showed higher response among the following doctors:

- those who are age 65 or above, attended a U.S. or Canadian medical school and who are located in an MSA with a population of at least 250,000 (odds ratio 4.2);
- those who were born in the USA or Canada and whose percent of hours worked in a hospital is zero or unknown (odds ratio 3.6);
- those who are pediatricians (odds ratio 2.8); are specialists (odds ratio 1.2); received a survey incentive of $75 with full protocol and were in the "remaining" (post-experiment) sample (odds ratio 1.2);
- and those whose present employment is "other" (odds ratio 1.1).

E. FINAL COMPUTATION OF THE WEIGHTS

The objectives when computing the national weights are (1) to minimize the risk of introducing bias on the sample estimates, and (2) to reduce the design effect of the sample estimates. Thus, after applying the non-response adjustments, post-stratification is necessary to match the adjusted weights to the population totals in the frame. The post-stratified weights are checked to see if trimming was needed to avoid extreme weight values.

After applying the adjustments to the weights for non-locatable physicians and for non-response among located physicians, the weighted counts for physicians who completed the

interviews or who were ineligible did not reproduce the frame totals for region and specialty.[13] Therefore, we formed 40 cells (10 regions by four specialty groups) and computed a ratio-type adjustment so that the sum of the non-response adjusted weights matched the frame counts for those cells. These adjustments were the frame count for a group divided by the corresponding sum of the non-response adjusted weights for the completed and ineligible interviews in the group.

After the post-stratified weights were developed, we examined the distribution to see if trimming was needed to address the potential of extreme weights that inflate the sampling variance of survey estimates. We concluded that trimming was not necessary. The sum of weights was 411,784.

[13] Specialists, plus the three subgroups of PCPs: general/family practice, internal medicine, and pediatrics.

References

CyBulski, K., M. Sinclair, F.B. Potter, and A.B. Ciemnecki. "Adjusting for Nonresponse Among Medicaid Households that Could Not Be Located or Were Located but Did Not Participate in the Minnesota Managed Care Survey." Paper presented at the International Conference on Survey Nonresponse, Portland, Oregon, 1999.

Folsom, R.E., and M.B. Witt. "Testing a New Attrition Nonresponse Adjustment Method for SIPP." Proceedings of the American Statistical Association, Survey Research Methods Section, 1994, pp. 428-433.

Hosmer, D.W., & Lemeshow, S. (2000). *Applied logistic regression (2nd Edition).* New York: Wiley

Iannacchione, V.G., J.G. Milne, and R.E. Folsom. "Response Probability Weight Adjustment Using Logistic Regression." Proceedings of the American Statistical Association, Section on Survey Research Methods, 1991, pp. 637-642.

Little, Roderick J.A. "Survey Nonresponse Adjustments for Estimates of Means." International Statistical Review, vol. 54, no. 2, 1986, pp. 139-157.

Potter, F.G., V.G. Iannacchione, W.D. Mosher, R.E. Mason, and J.D. Kavee. "Sample Design, Sampling Weights, Imputation, and Variance Estimation in the 1995 National Survey of Family Growth." Vital Health Statistics, series 2, no. 124, 1998.